



Robust rolling stock in rapid transit networks

Luis Cadarso, Ángel Marín *

Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Aeronáuticos, Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain

ARTICLE INFO

Available online 19 November 2010

Keywords:

Rapid transit
Robust rolling stock
Shunting

ABSTRACT

This paper focuses on the railway rolling stock circulation problem in rapid transit networks, in which frequencies are high and distances are relatively short. Although the distances are not very large, service times are high due to the large number of intermediate stops required to allow proper passenger flow. The main complicating issue is the fact that the available capacity at depot stations is very low, and both capacity and rolling stock are shared between different train lines. This forces the introduction of empty train movements and rotation maneuvers, to ensure sufficient station capacity and rolling stock availability.

However, these shunting operations may sometimes be difficult to perform and can easily malfunction, causing localized incidents that could propagate throughout the entire network due to cascading effects. This type of operation will be penalized with the goal of selectively avoiding them and ameliorating their high malfunction probabilities. Critic trains, defined as train services that come through stations that have a large number of passengers arriving at the platform during rush hours, are also introduced.

We illustrate our model using computational experiments drawn from RENFE (the main Spanish operator of suburban passenger trains) in Madrid, Spain. The results of the model, achieved in approximately 1 min, have been received positively by RENFE planners.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The preservation of historical centers and the corresponding social life, along with increased passenger demand in urban and suburban areas and the resulting traffic congestion in the central areas of cities, have led many cities to build or extend rapid transit systems. The Rolling Stock (RS) problem studies the assignment of a given train fleet in a dense urban network to optimally satisfy a given timetable and passenger demand.

The RS problem, at the strategic level, studies the fleet investments required to satisfy the demand pattern by making assumptions about the data over several years. Decisions to be made involve the most convenient vehicle fleet to buy, assuming a given service infrastructure. Thus, at the strategic level, infrastructure decisions are also made after considering the massive future use by the RS. Meanwhile, at a tactical level, the RS problem assigns the fleet to satisfy medium-level demand over a planning period of weeks or months. However, this study does not include changes in train composition or train maneuvers at the depots, that is, shunting operations are not considered in detail during this planning phase.

In a daily planning period, the data and the decisions must be considered in the context of a space-time network. A known demand and timetable are met by a given fleet. The RS model makes decisions about the aggregation and disaggregation of the different RS in the depot stations; in this phase, shunting operations are studied in detail and generic plans (the result of tactical planning) may be adjusted to meet the specific demands of particular scenarios. The problem can be stated as follows in the context of metropolitan rapid transit networks: given the train services' departure and arrival times and the expected numbers of passengers at each arc and in each period, and considering composition changes, find the optimal assignment of the RS to the train services.

Major complicating issues are the available shunting capacities at depot stations and RS sharing between lines. Shunting is related to the need for RS to be parked in shunting areas when the RS is not used for traffic during off-peak hours and for those maneuvers to match compositions during times between the beginning and end of the planning period. The shunting process is very complex for urban and suburban depots because depot stations are shared between RS moving on different train lines. This implies that depot station capacities may change in different time periods. This forces us to combine and split convoys to form trains and to consider the logistics of empty train movements in order to meet depot station capacities. However, empty trains will also be moved to ensure RS availability because RS is a very limited resource during rush hours, when the passenger demand is very high.

* Corresponding author. Tel.: +34 91 3366323; fax: +34 91 3366324.

E-mail addresses: luis.cadarso@upm.es (L. Cadarso),
angel.marin@upm.es (Á. Marín).

As we have stated above, both RS resources and depot station capacities are limited. Thus, we introduce composition changes to improve the availability of RS. Another complicating issue is rotation times. Rotations are the maneuvers performed at depot stations to change the direction of motion of the RS. It is assumed that the service time is the actual service time plus the rotation time; this time is known in the literature as the availability time. However, in rapid transit networks, in which capacities are limited and frequencies are on the order of the rotation time, we need to account for the depot station capacity in each period to avoid exceeding capacity.

For passenger demand, we use the expected number of passengers using each service given by RENFE. The expectation is based on historical data from the autumn of 2008. The RS model (RSM) assumes a certain flexibility in the passenger capacity of the trains as it attempts to provide comfortable service to passengers while also efficiently using RS.

This paper presents a specialized RSM describing rapid transit networks. Underground and suburban train problems are known as high-density network problems, in which the distances between the nodes are relatively short and the frequencies are high. The RSM will consider the optimization of train services' compositions, empty trains and the optimal management of convoys in the depot stations, all while considering the character and capacities of these types of RS and depots. The RSM is a first approach to the new subject of urban rapid transit networks, which, have been manually planned to date.

As stated above, empty movements and shunting operations are necessary in rapid transit networks. However, these operations may sometimes be difficult to execute and they can easily malfunction, causing localized incidents that could propagate through the entire network due to cascading effects. This is the case for composition changes. These operations will be penalized to selectively avoid them and their high malfunction probabilities. Alternatively, we can introduce robustness by avoiding empty train movements. During rush hours, the network is very congested and an empty movement using the same infrastructure as commercial trains increases the probability of an incident in the network. Empty movements at rush hours will also be penalized. Similarly, during rush hours, the number of passengers arriving at stations' platforms may be large in a given time period. Therefore, to avoid exceeding capacity due to train service delays, critic train services will be introduced to allow for greater capacities.

This paper is organized as follows. A literature overview is given in Section 2. We describe the RS problem for rapid transit networks in Section 3. In Section 4, the mathematical formulation is presented in detail. Section 5 contains the computational results based on a realistic case provided by RENFE. Finally, we present our conclusions in Section 6.

2. State of the art

Schrijver [1] minimizes the number of train units (self propelled train composed of carriages and locomotives) that must be employed in a line to avoid standing passengers. Although a unique type of train unit is considered, two subtypes can be defined. There is only one constraint for composition changes: between two consecutive trips, a change is allowed if the required vehicles are in the right place at the right time. An integer programming model is considered by Alfieri et al. [2] to determine the RS circulation for multiple RS types on a single line and on a single day, taking into account the order of the units in the train compositions. They use the concept of a transition graph to deal with this aspect. This concept is based on the assumption that for each trip, the next trip is known a priori. The problem is an integer multicommodity flow

problem, in which a feasible path in the transition graph is to be found simultaneously for each train. The objective is to minimize the number of units or the carriage-kilometers such that the given passenger demand is satisfied. The solution method involves decomposing the problem into subproblems and using their solutions to reduce the size of the original problem such that it becomes tractable using a commercial MIP solver. The approach is tested on real-life examples from NS, the main operator of passenger trains in the Netherlands. The model described by Alfieri et al. [2] was extended by Fioole et al. [3], to include combining and splitting trains, as happens at several locations in the Dutch timetable. They use an extended set of variables to locally obtain an improved description of the convex hull of the integer solutions. This method appears to substantially improve the lower bounds. Robustness is considered by counting the number of composition changes. Maróti [4] focuses on planning problems that arise at NS. He identifies tactical, operational and short-term rolling stock planning problems and develops operations research models for describing them. Then, he analyzes the considered models, investigates their computational complexity and proposes solution methods. The allocation of RS units to French TGV trains is studied by Ben-Khedher et al. [5]. The RS circulation must be adjusted to the latest demand known from the seat reservation system. Therefore, this problem contains a strong re-scheduling component. The objective is to maximize the expected profit for the company. To reach that goal, various operations research techniques are applied, including stochastic optimization, branch-and-bound, and column generation. A locomotive and carriage assignment problem was presented by Cordeau et al. [6]. The authors formulate the problem as a large integer program and use Benders decomposition to solve it. Computational experiments show that optimal solutions can be found quickly. In a subsequent paper by Cordeau et al. [7], their model was extended by considering various aspects such as maintenance of the RS. They propose a heuristic branch-and-bound algorithm for the extended model, solving the linear programming relaxations by column generation. A RS circulation problem related to the circulation of ICE train units in the German ICE network was described by Mellouli and Suhl [8]. In this case, the required capacities of the trains are known a priori. Carriages and locomotives first have to be combined into train units of certain pre-specified groups, and these train units then have to be routed through the network in an optimal way. The problem is modeled as an integer multicommodity flow problem on a multiple-layered network. Marín and Cadarso [9] define a model to study suburban rapid transit RS with convoys formed by three cars of the same type. The trains may be composed of one or two convoys in a dense network to attend to asymmetric demand and scheduling.

The problem of routing railway carriages through a railway network was considered by Brucker et al. [10]. The carriages should be used in timetable services or empty trains such that each timetable service can be operated with at least a given number of carriages, thereby satisfying passenger demand. Determining appropriate empty train movements increases the complexity of the problem. The objective is to minimize a non-linear cost function. The solution approach is based on local search techniques such as simulated annealing. Nielsen et al. [11] use a two-step model: once the timetable is known, they first generate circulations. That is, types of RS units are assigned to different trains and the authors then generate RS duties, assigning each RS unit to a determined number of tasks. The authors generate the duties externally to the model, identifying chains of tasks performed by the same type of RS. They then minimize carriage kilometers. Marín and Cadarso [12] present a model to study the Rapid Transit Routing problem for dense metropolitan networks. The RS does not identify specific trains to assign for each period and operation, but the train routing problem determines the operations to be rolled by

the individual convoys. The model has been tested in RENFE's suburban rail line. Cadarso and Marín [13] extend the model presented by Marín and Cadarso [12]. The authors develop a robust model that attempts to minimize the delay propagation in each sequence as well as the crew requirements at depot stations.

An elaborate introduction to the shunting problem including a solution approach and computational results, can be found in Freling et al. [14] and Lentink [15]. In these papers, the matching and parking subproblems are solved separately, resulting in solutions of sufficient quality. The matching problem is solved by a MIP model, and the parking problem is solved by column generation. Here, each column represents a subset of RS units that can be allocated to a single track in a feasible way. The problem of routing RS units between the platform and shunting areas of a station was studied by Van den Broek et al. [16]. They describe a capacity test that can be used to determine whether or not the proposed shunting movements fit between the regular train movements that had been planned during an earlier stage of the planning process. They test their model on several examples from NS.

In case of a disruption in the railway system, the first dispatching task is to keep the railway system running. These first decisions are taken under extreme time pressure. Therefore, decision proposals should be generated quickly. Timetable services must be provided with rolling stock of any type. A heuristic approach to rebuilding a passenger transportation plan in real time was proposed by De Almeida et al. [17]. This approach is intended for the management of major disruptions in which track capacity is greatly reduced. After computing an initial score for each train, a greedy algorithm is run to select trains. It is assumed that train compositions are not changed throughout the overall disturbance. However, a drawback of this approach is that it is global in nature and may impact parts of the transportation plan that were not disrupted. This makes the solutions more difficult for the users to understand.

Liebchen et al. [18] introduced the concept of the Price of Recoverability as a generic framework for modeling robustness issues in railway scheduling problems. However, the notion is theoretical in nature and is not straightforward to use in specific problems. Cacchiani et al. [19] explore the application of the Price of Recoverability idea to railway RS planning. They are particularly interested in practically computable recoverability measures, thereby evaluating the robustness of real-life RS schedules. They propose lower bounds that can be computed by solving mathematical programs. Their focus lies in real-life resource scheduling problems that are formulated as mathematical programs. Cacchiani et al. [20] explore the possibility of applying the notions of Recoverable Robustness and Price of Recoverability to railway RS planning, being interested in recoverability measures that can be computed in practice, thereby evaluating the robustness of RS schedules. They evaluate the approach on real-life RS planning problems encountered by NS.

The RS problem is very similar to the fleet assignment problem in the airline industry. Different aspects of this problem have been deeply researched. Daskin and Panayotopoulos [21] present an integer program that assigns aircraft to routes (which they define as sequences of flight legs originating and terminating at the same airport) in single-hub networks. Lagrangian relaxation is used to find an upper bound, and heuristics are used to find specific solutions. Abara [22] presents a model that can be used in more general airline networks, but the model has some limitations due to the use of connection arcs as decision variables. The model size explodes unless limits are placed on the connection opportunities. Another limitation is that different flying times and turn times (minimum ground service times) are not allowed for different fleets. Hane et al. [23] present a multicommodity flow model.

They show a number of ways to reduce the problem size: variable aggregation, cost perturbations, dual simplex with steepest-edge pricing, and intelligent branch and bound strategies. Fleet assignment models have been widely applied in practice and costs have been significantly reduced as we can see in Delta Airlines [24].

2.1. Contributions

In the mentioned literature, empty movements and rotations have not been considered. However, they are very important for our problem. As we are studying rapid transit networks with limited resources and high frequencies, we have to account for these possibilities. Empty train movements that do not attend to passenger demand are modeled using a new variable. Changing the direction of movement of the trains is usually implicit in the trip time, but we use a new variable to include it explicitly.

As we are studying rapid transit networks, capacity is not fixed because standing passengers can exist in different configurations. Therefore, we must include these different capacity configurations in the model formulation.

Most of the contributions mentioned above assume that the sequences of train services or a set of possible sequences are known. However, in our approach we consider a multicommodity flow model in which commodity flows are studied. Once these flows are known, train sequences will be determined in the routing problem.

3. The rolling stock problem in rapid transit networks

In this section, the RS problem is described in detail. First, the rapid transit network is introduced. Next, we describe train services and shunting in rapid transit networks. Then, we introduce passenger demand, and finally, we establish the concept of robustness in the RS problem.

3.1. The rapid transit network

As mentioned above, rapid transit networks are characterized by high frequencies and short distances. However, for the Madrid case study (Fig. 1), service times are large even though distances are short. This is due to the large number of intermediate stops required to meet passenger demand.

In this network we can distinguish two different principal types of stations represented by $s \in S$. The first type is characterized by train services that only attend to passenger demand. The second type is called a depot station. In these stations, shunting operations can also be performed. That is, attached to the passenger station is a depot where trains are driven to be parked or shunted. Depot stations are represented by $s \in S_d \subset S$. Some of the depot stations are specially characterized because they share their capacity among different lines. Because of this capacity sharing, their capacity may change during different time periods, complicating shunting operations. This time-dependent capacity is determined in a previous planning phase jointly with the infrastructure manager.

The existing infrastructure linking different stations is represented by arcs, $a \in A$. Between two stations, two different arcs exist, one for each direction of movement. Therefore, every arc a is defined by its departure and arrival station and by its length (e.g., in kilometers).

The planning time is discretized into time periods, $t \in T$. Due to the high train frequencies, the duration of one time period is set to 1 min. The existing physical network is replicated once for each time period existing in the planning period (e.g., 20 h).

In this section, we have introduced the space-time network on which we will develop our model formulation.



Fig. 1. RENFE Cercanías Madrid Rapid Transit Network.

3.2. Train services

Once the space-time network has been defined, train services within the network are known. Each train service is represented by $l \in L$. Train services are defined as commercial trains operating in the network to meet passenger demand. They are characterized by their departure depot station; their arrival depot station; every arc they travel on, defined by $a \in A | C \subset A$; and their departure time. The length of a train service l is the sum of the lengths of the arcs used by the train service. Consequently, an indicator $\alpha_{l,s,t}$ containing the timetable is known. It will take the value -1 (1) if train service l departs (arrives) from (at) depot station s during time period t .

Each train service l will have to be assigned a RS material (i.e., type of train) $m \in M$. In the rapid transit network of Fig. 1, material is not mixed within the same train service, i.e., the same material is used for each train service. However, different services may use different materials.

As mentioned before, different types of carriages and locomotives (which can move themselves) exist, and naturally, a carriage cannot move without a locomotive. To enable train movement, carriages are attached to locomotives of the same material. When locomotives are attached to both sides of the carriage, a convoy is obtained. A convoy can move in both directions. The composition $c \in C$ assigned to a train service will be given by the number of convoys forming the train service, and convoys are always from the same material. A train type and composition is given by $m, c \in M, C$.

3.3. Shunting

As we have mentioned above, shunting operations complicate rapid transit networks because the performance time is on the order of the service frequency time. This fact mandates that we explicitly include the rotation time. In the literature, rotations have usually been represented in the available time, which is the travel time plus the rotation time. With this assumption, the rolling

material is made available at the arrival station after the rotation has been performed, without taking the rolling material into account when calculating the depot station capacity constraints. However, rotation maneuvers are now included in the model through a new variable, to account for the high frequencies and the lack of capacity in depot stations. Thus, every departing train service must have time to perform a rotation. That is, the material assigned to a train service must be available at the depot station a certain number of periods (rotation time periods) before the departure time. Hence, we assume that every train service must change its direction of movement before its departure time.

Composition changes are the other shunting operations performed at depot stations. Convoys of the same type can be aggregated or disaggregated. The time needed to perform a composition change is greater than that needed for rotations. The rotation time (i.e., the time needed to enable the train to depart) is included in these operations. Moreover, when a composition change is performed, the number of convoys forming a train changes: it may increase or decrease. For this reason, two different variables are included in the model: one to represent aggregation and another to represent disaggregation. Although composition changes enable the network operator to use smaller fleet sizes, it is always a complicating operation, due to the necessity of human resources and the possibility of failure in the mechanical system governing the process.

As mentioned above, the network is characterized by high frequencies and a lack of capacity in depot stations. These facts make it difficult to operate the network without empty movements. These are defined by an origin, a destination and a departure time. Empty movement time is the same as that of train services between the same stations because the same infrastructure is used by both trains. Empty movements can help satisfy both capacity and rolling stock material availability in depot stations.

3.4. Passenger demand

The passenger demand for this problem is treated as a passenger flow $g_{a,l}$ through each arc a belonging to each train service l . This passenger flow is obtained from historical data under normal conditions (i.e., assuming that the train services matched the designed timetable). In this case, the passenger flow $g_{a,l}$ is known and is used to fix the required capacity for each arc in each service, assuming that the passenger flow carries complete information about the train service timetable and the capacity that will be offered (i.e., assuming equilibrium has been reached).

Under the above hypothesis, the model will treat the passengers from a centralized point of view (i.e., only the operator criteria are optimized). However, since the proposed problem relates to a suburban rapid transit network, it is obvious that every passenger will have the option to choose any other available company or transportation mode. Thus, the operator has to factor in passenger behavior to avoid losing passengers to other transportation companies.

The first and best choice would be to include the actual passenger behavior. Although this approach would give almost the true optimum, it is very complicated to include it in the model in a realistic way due to the large number of possibilities involved in rapid transit networks and their competitive operators. In a simplified approach, passenger behavior can be summarized as follows: if the passenger maintains his/her satisfaction with the transportation mode, he/she will remain in the system. As long as the system operator maintains certain standards within the transportation system, we assume that the passenger flow is known (i.e., the equilibrium mentioned above is achieved).

Under the assumption that public timetables are met, transportation standards might be described by the capacity offered in

each train service. The operator will always try to offer comfortable capacity in each train service. It is obvious that when more capacity is offered, more passengers will travel comfortably. However, offering more capacity increases operating costs dramatically. Therefore, the composition assigned to each train service will be a tradeoff between the operating costs and the behavior of passengers (represented by their comfort level).

For each convoy formed by one material type m , the passenger capacity is known. There is a fixed seating passenger capacity and a variable standing passenger capacity. Multiple possibilities arise when considering the standing passenger capacity. The aim is to obtain adequate passenger capacity for every train service. This may be obtained with different configurations for standing passengers. We define comfortable capacity as the full seating capacity and fewer than 3 pax/m² standing. If this capacity is exceeded, the passengers above this level are deemed passengers in excess. If the density of standing passengers is between 3 and 4 pax/m², each passenger in excess represents a moderate penalty, because the operator would like to obtain 3.5 pax/m². If the density of standing passengers exceeds 4 pax/m², passengers in excess are highly penalized because this situation is deemed very uncomfortable.

The latter can be formulated as a piecewise penalty function whose breakpoints will depend on the composition assigned to each train service. This is because the passengers admitted in each piece depend on the composition assigned, which is a variable in the problem.

The demand is not symmetric and can vary greatly from one arc to the next. For this reason, we introduce critical arcs to make the problem more tractable. These critical arcs are the arcs with maximum passenger load and will determine the employed train composition. If an arc is not considered critical, then it will not be included in the problem. To realistically represent passengers in excess for each train service, critical arcs are chosen in the following manner: every critical arc in the line during the day (i.e., every arc with maximum passenger load on each train service) is chosen as a critical arc for every train service. In the remainder of the paper, arcs refer to critical arcs only.

3.5. Robustness

As mentioned before, robustness is introduced through composition changes and empty movements.

When a composition change is performed, multiple failures can occur, forcing the train to be parked for a long time and causing an incident. The mechanical system used to perform a composition change is automatic, but it often fails and requires extra time to enact the change. Moreover, during composition changes, the brakes' pneumatic circuit must be joined or separated depending on the performed operation. This is always a difficult and complicating issue, and human resources are required to perform it. Above all, composition change times are overestimated to account for the effects mentioned above and to try to introduce robustness into the system. Finally, if a malfunction has occurred it must be contained to avoid cascading effects. Containment of cascading effects is easier if the incident occurs during off-peak hours when more RS material is available at the depot station and other material can depart from the station to attend to as much of the demand as possible. In our model, this is treated by harshly penalizing composition changes during rush hours (see coefficient $\theta_{s,t}$, which depends on the station s and the time period t).

Similarly, empty movements during rush hours complicate network operation because they use the same infrastructure as commercial train services. In addition, they obviously require human resources. Although human resources are always available at depot stations to perform composition changes and empty

movements, it is better to keep these resources in the depot station to alleviate possible incidents during rush hours. Therefore, empty movements during rush hours are also heavily penalized. It is also better to avoid (if possible) empty movements to destination depot stations with time-dependent capacities (i.e., stations that are shared with different lines). This idea is represented by the $\theta_{s,s',t}$ coefficient, which penalizes empty movements between depot stations s, s' within departure time period t .

Another aspect that could be interpreted as robustness is the critical train. During rush hours, the number of passengers arriving at station platforms per unit time period is very large. This implies that a small delay at departure time will change the actual passenger flow. A train service is considered a critical train if it comes through stations that have a large number of passengers arriving at the platform during rush hours. This is reflected in our model through a greater penalty per passenger in excess for these train services, according to the operator's wishes.

Finally, the system is made more robust by assigning only one material type per line (i.e., for every train service operating the same line, the material must be equal). This constraint allows for all material on one line to be swapped between different train services at depot stations serving that line. Thus, there will be more opportunities to swap train services if an incident occurs, and the propagation of the incident can be mitigated easily. Moreover, at shared depot stations, there could be as many material types as there are lines using the station, but it will be shown that some lines share the same material type and depot stations so that the lines can interchange RS material.

4. The rolling stock model

In the RS model (RSM), the types and compositions of the train services are determined. The convoys are usually formed by two locomotives and one carriage of the same type. Convoys from different types are not compatible and cannot be mixed in the same train. For the Madrid case study (Fig. 1), a train may be formed of one or two convoys. Changes in composition will only take place in the initial or final depot stations of the train service. The distribution of convoys in the depot stations will be the same at the beginning and end of the planning period.

Light maintenance requirements are not included in the model because they are met during off-peak hours. Once the RS has been assigned, the train routing, which we refer to as train sequences, has to be determined (see Cadarso and Marín [13]). When train sequences are designed, the location of all RS material in the space-time network will be known. Thus, material that requires light maintenance is assigned to a sequence containing light maintenance opportunities during off-peak hours. For heavy maintenance, the fleet size is supposed to be large enough to take the RS material requiring heavy maintenance out of the rapid transit network.

We include material capacity constraints in depot stations. We could also include capacity constraints for sections of the network. However, the scheduling of train services has been properly designed in timetable planning: the capacity of every section is considered and adequate supervision of the infrastructure manager exists in this planning phase. Thus, section capacity constraints are not necessary because they are automatically matched. Likewise, train services must not be longer than station platforms: for each train service we consider the shortest platform in its path, so we only include one constraint per train service for the platform length requirement.

In our model, the relationships between the data and variables are considered within a directed space-time graph, $G(S,A)$, where S is the set of stations and A is the set of arcs. Each arc a is defined by

(s, t, s', t') , where s and s' are the origin and destination nodes, t is the departure time, and t' is the arrival time. That is, $t' = t + t_a$, where t_a is the time to move from s to s' . It is assumed that this time is known and fixed for each arc. This means that in the RSM, in which an arc is denoted by a , this may be understood as $a = (s, s', t)$.

In this space-time graph, the train services are given. These services are a known sequence of sections, with a known departure time from the first station. We define the train set as the pair (l, t) , where l is the train service number and t is the departure time period, implicit in the train service number.

The RSM arises as an extension of the model proposed in Marín and Cadarso [9]. In the model presented in this paper, special attention is given to shunting in depot stations and robustness. The RSM mathematical formulation follows:

Sets:

- $L(l)$: the set of train services. Each train service is characterized by an origin, a destination and a departure time.
- $T(t)$: the set of time periods.
- $S(s)$: the set of stations.
- $A(a)$: the set of arcs.
- $M(m)$: the set of convoy types.
- $C(c)$: the set of convoy numbers. The index of this set is the number of convoys composing the train.
- $Al(a, l) = 1$ if arc a is used by the train service l , and $= 0$ otherwise.
- $Sl(s, l) = 1$ if station s has the minimum platform length in service l , and $= 0$ otherwise.
- $Sc(s)$: the set of depot stations.
- $CT(t)$: the set of count time.

Parameters:

- $c_{m,c}$: operating cost per rolled kilometer of convoy type m using c convoys.
- ic_m : investment cost of convoy type m . This parameter may represent a leasing cost for material from other lines.
- $p_{a,l}^{3-4}$: penalty per passenger in excess between 3 and 4 pax/m² in arc a and train service l .
- $p_{a,l}^{4-10}$: penalty per passenger in excess between 4 and 10 pax/m² in arc a and train service l .
- $\theta_{s,s',t}$: penalty for empty movement between depot stations s, s' with departure time period t .
- $\vartheta_{s,t}$: cost of composition change in depot station s and in time period t .
- $\alpha_{l,s,t} = -1$, if train service l leaves from station s in time period t ; $= 1$, if train service l arrives at station s in time period t ; $= 0$, otherwise.
- χ_m : fleet size for convoys of type m .
- $g_{a,l}$: expected passenger flow in arc a used by train service l .
- q_m^3 : passenger capacity (seating+standing) for the 3 pax/m² configuration in convoys of type m .
- q_m^4 : passenger capacity (seating+standing) for the 4 pax/m² configuration in convoys of type m .
- q_m^{10} : passenger capacity (seating+standing) for the 10 pax/m² configuration in convoys of type m . This number is large to avoid infeasible solutions.
- $cap_{s,t}$: time-dependent capacity of station s in time period t .
- l_m : length of a convoy of type m .
- pl_s : the platform length for each station s .
- o_c : the ordinal of c .
- r_s : the rotation time duration in depot station s .
- e_s : the junction time duration in depot station s .
- d_s : the separation time duration in depot station s .
- t_i, t_f : the initial and final times of the planning period.
- Km_l : the number of kilometers rolled by train service l .
- $et_{s,s'}$: empty movement time from station s to station s' .

- $\beta_{l,t,t'} = 1$, if train service l that departed during time period t' is still rolling during time period t ; $= 0$, otherwise.
- $\gamma_{s,t,t'} = 1$, if a rotation that finishes at depot station s during time period t' was being performed during time period t ; $= 0$, otherwise.
- $\mu_{s,t,t'} = 1$, if a composition change that started at depot station s during time period t' is still performing during time period t ; $= 0$, otherwise.
- $\xi_{s,s',t,t'} = 1$, if an empty movement between depot stations s and s' that departed during time period t' is still rolling during time period t ; $= 0$, otherwise.

Variables:

- $x_{l,m,c} = 1$, if train service l uses convoy type and composition (m, c) ; $= 0$, otherwise.
- $em_{s,s',t}^{m,c} = 1$, if empty movement from s to s' begins at depot station s during period t with type and composition (m, c) ; $= 0$, otherwise.
- $yt_{s,t}^{m,c}$: an integer variable, the number of trains with composition of type (m, c) , in station s during time period t (train inventory in station s).
- yn_m : an integer variable, the number of convoy type m to buy. Used to avoid infeasibilities in the model. It may also represent leasing costs from other lines.
- $\pi_{a,l}^{3-4}$: a positive variable, the number of passengers in excess between 3 and 4 pax/m² that use the train service l at arc a .
- $\pi_{a,l}^{4-10}$: a positive variable, the number of passengers in excess between 4 and 10 pax/m² that use the train set l at arc a .
- $cc_{s,t}^{m,c}$: an integer variable that counts composition changes performed at depot station s during period t from type and composition (m, c) .
- $e_{s,t}^{m,c} = 1$, if we begin the junction at depot station s during period t with type and composition (m, c) ; $= 0$, otherwise.
- $\delta e_{s,t}^{m,c} = 1$, if we begin the separation at depot station s during period t with type and composition (m, c) ; $= 0$, otherwise.
- $\rho_{s,t}^{m,c} = 1$, if a rotation is finished at depot station s during period t with type and composition (m, c) ; $= 0$, otherwise.

The RSM for rapid transit networks is formulated as a multi-commodity flow model. Some new aspects are contributed in the presented formulation: we have introduced multiple passenger capacities in the model depending on the configuration of standing passengers, using a piecewise formulation. In multicommodity flow balance constraints, we have included the material type and the composition, which can be understood as a subtype; composition changes have been included, which interchange commodities of different subtypes; similarly, rotations and empty movements have been included in these constraints. Moreover, a new group of constraints has been included to ensure that departing material has performed the mandatory shunting operations.

4.1. Objective function

$$\begin{aligned} \min z = & \sum_{l \in L} \sum_{m \in M} \sum_{c \in C} c_{m,c} \cdot km_l \cdot x_{l,m,c} \\ & + \sum_{s,s' \in Sc} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \theta_{s,s',t} \cdot c_{m,c} \cdot km_{s,s'} \cdot em_{s,s',t}^{m,c} \\ & + \sum_{s \in Sc} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \vartheta_{s,t} \cdot cc_{s,t}^{m,c} + \sum_{m \in M} ic_m \cdot yn_m \\ & + \sum_{l \in L} \sum_{a \in Al} p_{a,l}^{3-4} \cdot \pi_{a,l}^{3-4} + \sum_{l \in L} \sum_{a \in Al} p_{a,l}^{4-10} \cdot \pi_{a,l}^{4-10} \end{aligned}$$

In the objective function, a number of different costs are minimized. First, the operating costs of commercial train services are minimized. In the second term, the operating costs of empty movements are also minimized (cost per rolled kilometer equal to that of commercial

service trains). However, the coefficient $\theta_{s,s',t}$ increases operating costs for some empty movements to introduce robustness to the system. Another shunting cost, related to composition changes, is minimized in the third term. Through coefficient $\vartheta_{s,t}$, the cost of composition change is made dependent on station and time period. Robustness is introduced by minimizing composition changes, as these changes usually malfunction. A special cost is introduced in the fourth term to account for the possibility of leasing material from other lines. For computational purposes, this is equivalent to an infinite cost to avoid infeasibilities in the model. Finally, costs related to passengers in excess are introduced. The first cost appears if the density of standing passengers is between 3 pax/m² and 4 pax/m². Another cost is then introduced for standing passengers between 4 pax/m² and 10 pax/m². Both terms contribute to minimize the number of excess passengers.

Decision variables are subject to the constraints described in the following subsections.

4.2. Train service constraints

$$\sum_{m \in M} \sum_{c \in C} x_{l,m,c} = 1 \quad \forall l \in L \quad (1)$$

$$\sum_{m \in M} \sum_{c \in C} o_c \cdot l_m \cdot x_{l,m,c} \leq pl_s \quad \forall l \in L, s \in SI \quad (2)$$

Constraints (1) require that every train service is assigned a RS composition. Constraints (2) ensure that every train service matches the platform length requirements for its path.

4.3. Demand constraints

$$\sum_{m \in M} \sum_{c \in C} o_c \cdot q_m^3 \cdot x_{l,m,c} \geq g_{a,l} - \pi_{a,l}^{3-4} - \pi_{a,l}^{4-10} \quad \forall l \in L, a \in Al \quad (3)$$

$$\pi_{a,l}^{3-4} \leq \sum_{m \in M} \sum_{c \in C} o_c \cdot (q_m^4 - q_m^3) \cdot x_{l,m,c} \quad \forall l \in L, a \in Al \quad (4)$$

$$\pi_{a,l}^{4-10} \leq \sum_{m \in M} \sum_{c \in C} o_c \cdot (q_m^{10} - q_m^4) \cdot x_{l,m,c} \quad \forall l \in L, a \in Al \quad (5)$$

Constraints (3) require that the capacity assigned to each train service is sufficient to satisfy the passenger demand requirements. If the capacity is insufficient, the number of excess passengers is calculated. These passengers are limited in number by constraints (4) and (5), one for each group of passengers.

4.4. Material constraints

$$\begin{aligned} y_{s,t}^{m,c} + \sum_{l \in L} x_{l,m,c} + o_{c-1} \cdot e_{s,t}^{m,c-1} + o_{c+1} \cdot \delta e_{s,t}^{m,c+1} + \rho_{s,t}^{m,c} + \sum_{s' \in Sc} em_{s',t}^{m,c} \\ = y_{s,t}^{m,c} + \sum_{l \in L} x_{l,m,c} + o_{c+1} \cdot e_{s,t}^{m,c} + o_{c-1} \cdot \delta e_{s,t}^{m,c} + \rho_{s,t+r_s}^{m,c} \\ + \sum_{s' \in Sc} em_{s',t}^{m,c} \quad \forall s \in Sc, t \in T, m, c \in M, C \end{aligned} \quad (6)$$

$$\begin{aligned} \sum_{s \in Sc} \sum_{c \in C} o_c \cdot y_{s,t}^{m,c} + \sum_{l \in L} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \beta_{l,t',t} \cdot x_{l,m,c} \\ + \sum_{s \in Sc} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \gamma_{s,t',t} \cdot \rho_{s,t'}^{m,c} + \sum_{s,s' \in Sc} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \xi_{s,s',t',t} \cdot em_{s',t'}^{m,c} \\ \sum_{s \in Sc} \sum_{t' \in T} \sum_{c \in C} \mu_{s,t',t} (o_{c+1} \cdot e_{s,t'}^{m,c} + o_c \cdot \delta e_{s,t'}^{m,c}) \\ \leq \lambda_m + y_{n_m} \quad \forall m \in M, t \in CT \end{aligned} \quad (7)$$

Constraints (6) describe the balance of material flow in each depot station for every time period. Material parked in the immediately preceding period plus the material arriving by commercial train

services, empty movements and finished composition changes and rotations must be equal to the material parked in the next period plus the departing commercial train services, empty movements and composition changes and rotations that begin in the next period. Constraints (7) require that the fleet size is large enough to satisfy the network flows. This is only verified at one period because of the material flow constraints.

4.5. Shunting constraints

$$\sum_{\substack{l \in L \\ a_{l,s,t} = -1}} x_{l,m,c} = \rho_{s,t}^{m,c} + e_{s,t-e_s}^{m,c-1} + \delta e_{s,t-d_s}^{m,c+1} \quad \forall s \in Sc, t \in T, m, c \in M, C \quad (8)$$

$$\begin{aligned} \sum_{m,c \in M,C} o_c \cdot y_{s,t}^{m,c} + \sum_{t' \in T} \sum_{m,c \in M,C} \mu_{s,t',t} (o_{c+1} e_{s,t'}^{m,c} + o_c \delta e_{s,t'}^{m,c}) \\ + \sum_{t' \in T} \sum_{m,c \in M,C} o_c \cdot \gamma_{s,t',t} \cdot \rho_{s,t'}^{m,c} \leq cap_{s,t} \quad \forall s \in Sc, t \in T \end{aligned} \quad (9)$$

$$cc_{s,t}^{m,c} = e_{s,t}^{m,c} + \delta e_{s,t}^{m,c} \quad \forall s \in Sc, t \in T, m, c \in M, C \quad (10)$$

$$y_{s,t}^{m,c} = y_{s,t}^{m,c} \quad \forall s \in Sc, m \in M, c \in C \quad (11)$$

Constraints (8) require that every departing commercial train service has performed the necessary rotation. Composition changes are included because, as stated above, they also include the rotation time. Constraints (9) are depot station capacity constraints. The material in each depot station is accounted for in every period to avoid exceeding the capacity $cap_{s,t}$, which depends on time. Constraints (10) count every composition change into one single variable. Finally, constraints (11) ensure that the distribution of the fleet throughout the depot stations is equal at the beginning and end of the planning period.

4.6. Variable domain

$$x_{l,m,c} \in \{0,1\} \quad \forall l \in L, m \in M, c \in C \quad (12)$$

$$em_{s',t}^{m,c} \in \{0,1\} \quad \forall s,s' \in Sc, t \in T, m \in M, c \in C \quad (13)$$

$$e_{s,t}^{m,c} \in \{0,1\} \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (14)$$

$$\delta e_{s,t}^{m,c} \in \{0,1\} \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (15)$$

$$cc_{s,t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (16)$$

$$y_{s,t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (17)$$

$$y_{n_m} \in \mathbb{Z}^+ \quad \forall m \in M \quad (18)$$

$$\rho_{s,t}^{m,c} \in \mathbb{Z}^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (19)$$

$$\pi_{a,l}^{3-4} \in \mathbb{R}^+ \quad \forall l \in L, a \in Al \quad (20)$$

$$\pi_{a,l}^{4-10} \in \mathbb{R}^+ \quad \forall l \in L, a \in Al \quad (21)$$

In the solution approach, some variable domains can be relaxed. Continuity in the excess passenger variable is justified because the passenger flow value is an expectation and could be non-integer. Moreover, some variables in the model formulation are defined as a sum of binary and integer variables, and may therefore be relaxed. These variables are $y_{s,t}^{m,c}$, $cc_{s,t}^{m,c}$, $\rho_{s,t}^{m,c}$ and y_{n_m} . Thus, constraints (16)–(17)–(18)–(19) can be replaced by constraints (22)–(23)–(24)–(25), respectively.

$$cc_{s,t}^{m,c} \in \mathbb{R}^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (22)$$

$$yt_{s,t}^{m,c} \in R^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (23)$$

$$yn_m \in R^+ \quad \forall m \in M \quad (24)$$

$$\rho_{s,t}^{m,c} \in R^+ \quad \forall s \in Sc, t \in T, m \in M, c \in C \quad (25)$$

5. Computational experiments

All of our computational experience is for realistic cases drawn from RENFE's regional network in Madrid, also known as "Cercanías Madrid" (Fig. 1). This network is composed of 10 different lines with almost 100 stations. All data are from the year 2008. Approximately one million passengers use "Cercanías Madrid" every day.

The network presented in this study case (Fig. 1) is characterized by its modular structure. That is, in real-life it is separated into different and independent modules for operating purposes. Every module has its own infrastructure (stations, depot stations, sections, etc.). In this way, the RS material cannot be easily transferred from one module to another; it can be done, but extra human resources and time are needed to accomplish it; these extra resources are represented by the leasing cost in the objective function.

As mentioned above, some depot stations are shared between different lines. In these cases, we have used the total capacity of the station, and the model decides how to allocate the capacity to the material of different lines. However, some depot stations are shared between different modules. This is where the time-period-dependent capacity originates. The capacity assigned to each module is decided jointly with the infrastructure manager during timetable planning, and the capacity assigned to each module may change during the planning period.

From this point of view, it makes no sense to solve the whole network because we have all of the data for every module. Hence, two different cases (network modules) have been studied. The first case is line C5. This line can be considered an independent line for RS assignment purposes. However, it shares some depot stations. We have chosen this module because it has the highest frequency in the network. The second case consists of lines C3–C4. Although these are two different lines, they use the same material and share some depot stations; therefore, the two lines can interchange RS material.

The presented study cases have common depot stations. Thus, these modules are representative of the rest of the network, where there are more modules composed of one and two lines.

Our runs were performed on a Personal Computer with an Intel Core2 Quad Q9950 CPU at 2.83 GHz and 8 GB of RAM, running under Windows Vista 64 Bit, and our programs were implemented in GAMS/Cplex 11.1.

5.1. Study case 1: Line C5

Line C5 has more than 320 train services scheduled each day with frequencies on the order of 3 min at rush hour, equivalent to the rotation time in this line. The line has 22 stations (Fig. 2) and 4 depot stations: Mostoles el Soto, Atocha, Fuenlabrada and Humanes. There is one material type available, and the train

services can be of simple (one convoy) or double (two convoys) composition.

The convoys for the material in Line C5 have definite characteristics, as shown in Table 1. The train capacity is divided into seated and standing passengers. The number of seats is fixed, and a density value is defined for standing passengers. In every convoy, for example, for a density of 3 pax/m² we would have 240 fixed seats plus 261 standing passengers (i.e., 501 passengers per convoy). However, for a density of 4 pax/m², we would have the same 240 seats but a total capacity of 588 passengers.

In a daily planning period from 5:00 a.m. to 1:00 a.m. divided into one minute periods, we have 1140 time periods. The rotation time is 3 min at every depot station.

The RS MIP model size for this case is shown in Table 2. The RSM numbers of discrete and continuous variables, constraints and non-zero elements are given for the complete model (RSM) and for the reduced model (Reduced RSM) obtained using the Cplex presolver.

The primary model parameters for the operator are those penalizing excess passengers, ($p_{a,l}^{3-4}, p_{a,l}^{4-10}$) and the robustness parameters, ($\theta_{s,s',t}, \vartheta_{s,t}$).

Robustness parameters are obtained from operators. For example, in line C5, the Atocha depot station is shared among more than five different lines. This causes the capacity of C5 material to vary strongly during the planning period. Robustness parameters are chosen to try to avoid (if possible) composition changes and empty movements with the destination Atocha.

Once the robustness criteria are fixed, different solutions can be obtained by varying the penalties for excess passengers. These results are summarized in Table 3. For simplicity, we show three

Table 1
Convoy capacity and length.

Material type	Seats	Standing	Density (Pax/m ²)	Length (m)
m1	240	261	3	80
		348	4	
		870	10	

Table 2
Line C5 RS model: the number of variables, constraints and non-zeros.

	RSM	Reduced RSM
# Discrete variables	58628	47207
# Continuous variables	175924	11575
# Constraints	89317	19058
# Non-zero elements	797321	139388

Table 3
Line C5 RS model solutions.

$p_{a,l}^{3-4}, p_{a,l}^{4-10}$	#C	TSOC	EMOC	PEC	#CC	OI	ST
0.5,3	66	78331.44	1635.76	3019	16	43.35	13.05
1,5	64	80099.76	1265.04	3554	20	42.30	24.56
3,6	65	82896.24	1281.36	4403	18	40.73	8.55



Fig. 2. Line C5.

different cases in which the penalties for excess passengers are constant. However, these penalties are enlarged for critical train services to account for possible delays in train services. The pair of excess passenger penalties ($p_{a,t}^{3-4}, p_{a,t}^{4-10}$) is listed in the first column, the number of convoys used in the proposed solution (#C) in the second column, the train service operating costs (TSOC) in the third column, the empty movement operating costs (EMOC) in the fourth column, excess passenger costs (PEC) are shown in the fifth column, the number of composition changes (#CC) in the sixth column, the occupation index (OI) in the seventh column (obtained as an average value of every train service's OI), and the solver solution time (ST) in seconds in the final column.

An important cost to the operator is the maintenance cost. For example, for this material, there is a daily fixed cost of, say, 400 €. Given the importance of the number of convoys used in the network, different solutions must account for that number. If we compare any of the proposed solutions in Table 3 with the current solution provided by RENFE's operators in Table 4, we can see that the number of used convoys is slightly smaller. For different excess passenger penalties, we can see how the number of convoys used varies; this variation is due to the number of composition changes performed, to the fact that increased penalties increase the PEC and more capacity must be offered, to material flows in the network subject to capacities, etc. For example, in Table 3 we can see how the solution with more composition changes is the solution that uses fewer convoys.

Train service costs and empty movement costs are also reduced. However, excess passenger costs increase. This increase arises because if fewer convoys are used, less capacity is offered to the passengers. For this reason, the occupation index also increases, and composition changes appear to enable different compositions

of the train services. It is also worth noting that the solution time is nearly real time, substantially reducing the manual planning time.

As Atocha station is shared with many other lines and materials, there is no composition change or empty movement with a destination of Atocha in any of the proposed solutions. However, empty movements from Atocha to other depot stations appear to provide capacity to this congested station.

As we have stated above, robustness may be introduced with different approaches. Some of these approaches could prevent dangerous empty movements and composition changes. Dangerous empty movements and composition changes are recognized as those occurring during rush hours. To illustrate this effect, two different cases are shown. In the first case, no robustness is introduced (NoRob) (i.e., $\theta_{s,t} = 1$ and $\vartheta_{s,t}$ is equal to its nominal value for all possible cases). In the second case, robustness is introduced (Rob), increasing the values of these parameters at rush hours from 7:00 a.m. to 10:00 a.m. and from 14:00 p.m. to 17:00 p.m. The computational results are shown in Table 5. In both cases, the rest of the parameters are those used in the previous examples.

We can see the differences between both cases in Table 5. In the second column, the train service operating costs (TSOC) are shown. In the robust case, these costs are greater because more train services are set to double composition; this cost may represent the robustness cost. In the third column, the empty movement costs (EMOC) are shown, which stay equal to those of the no robustness case. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is reduced, achieving one of our goals. In a similar way, robustness is introduced in the solution by avoiding composition changes during rush hours (#CCRH). We note that for the cases studied, the passengers in excess costs (PEC) always remained similar or equal to those of the non-robust case. This is a desirable thing, because the price of robustness does not arise from passengers.

The network operator considers service to be of good quality when standing passengers do not exceed a density of 3.5 pax/m². We have therefore attempted to match this objective. Fig. 3 shows passengers exceeding a density of 3.5 pax/m² in RENFE's current solution for the entire C5 line in every time period.

Implicit in the parameter penalizing excess passengers, the critical train concept arises. That is, for arcs with great passenger flow at rush hours, these parameters are increased. For example, if under normal conditions we choose the pair (1,5) for these parameters, the pair would be (10,11) for train services coming through those arcs during rush hours.

We can see the results of our model in Fig. 4 for the case in which (1,5) represents the penalties for excess passengers. To show how critical trains work, we have enabled this feature for some train services in the morning rush hour. However, we have removed critical trains from the afternoon and evening rush hours.

Table 4

Line C5 RS current solution operated by RENFE.

	#C	TSOC	EMOC	PEC	#CC	OI
Current solution	74	109765.20	2232.12	874	0	28.30

Table 5

Comparing non-robust and robust solutions for line C5.

Case	TSOC	EMOC	#EMRH	PEC	#CC	#CCRH	OI	ST
NoRob	80099.76	1265.04	13	3554	20	4	42.30	24.56
Rob	80413.68	1265.04	10	3248	20	3	42.11	28.54

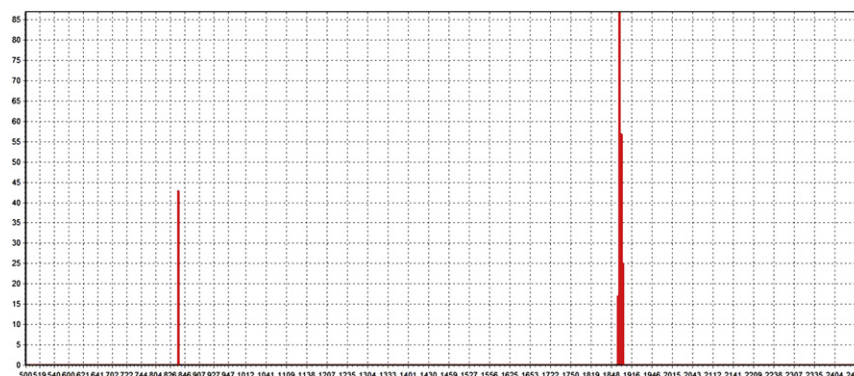


Fig. 3. The passengers in excess above 3.5 pax/m² configuration for line C5 in the RENFE solution.

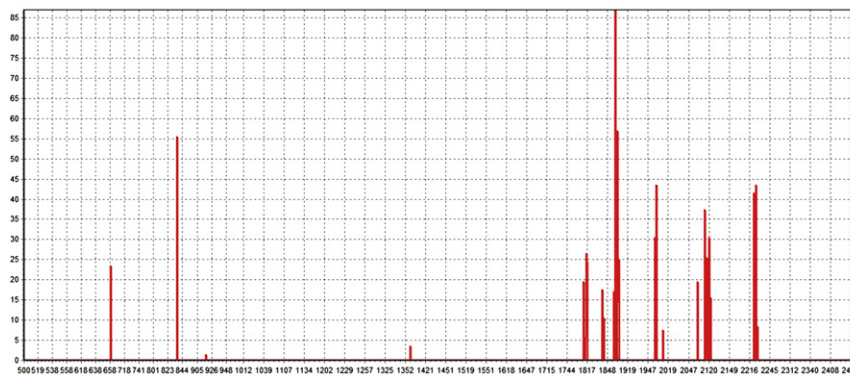


Fig. 4. The passengers in excess above 3.5 pax/m^2 configuration for line C5 in our proposed solution. Critical trains are enabled for the first rush hour but not for the following rush hours.

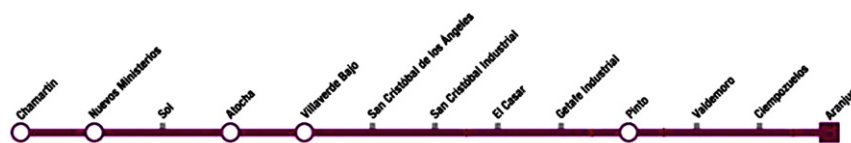


Fig. 5. Line C3.

Thus, during these two rush hours we do not consider possible delays in train services to cause consequent increases in the number of passengers at platforms. Train services will run with more passengers without accounting for critical trains, and they will not be able to absorb any extra demand due to delays. Moreover, quality of service during these rush hours will be worse. However, when using critical trains for the first rush hour in the morning, we can see how the quality of service is maintained at a level very similar to that of RENFE's current solution and, therefore, train services will be able to absorb the extra demand at the station platforms caused by small delays, providing robustness to the solution. We can appreciate the difference between using or not using critical trains in Fig. 4. Finally, we can state that the major difference between PEC in RENFE's current solution and our proposed model solution is due to passenger densities between 3 and 3.5 pax/m^2 , which are penalized as passengers in excess. We must mention that the results in Table 3 were obtained with critical trains enabled every rush hour.

5.2. Study case 2: Lines C3–C4

In lines C3–C4, there are nearly 400 scheduled train services each day with frequencies on the order of 10 min. In these lines, the rotation time is 2 min greater than that of line C5.

Line C3 is composed of 12 stations (Fig. 5) and 3 depot stations: Chamartín, Atocha and Aranjuez. In line C4, there are 18 stations (Fig. 6) and 8 depot stations: Parla, Parla Industrial, Getafe Centro, Atocha, Chamartín, Tres Cantos, Alcobendas and Colmenar Viejo. Some depot stations are shared between these lines.

There is one material type available, and the train services can have simple (one convoy) or double (two convoys) compositions. Because the same material is used for both lines, RS can be interchanged between them.

The convoys on lines C3–C4 have specific characteristics that are shown in Table 6. As shown, the train capacity is divided into seated and standing passengers. In every convoy we have a capacity of 277 seats, but the number of standing passengers depends on their



Fig. 6. Line C4.

Table 6
Convoy capacity and length.

Material type	Seats	Standing	Density (Pax/m ²)	Length (m)
m2	277	360	3	80
		480	4	
		870	10	

Table 7
Lines C3–C4 RS model: the number of variables, constraints and non-zeros.

	RSM	Reduced RSM
# Discrete variables	134168	97488
# Continuous variables	242256	19565
# Constraints	107189	29030
# Non-zero elements	1476245	267795

density, for example, for a density of 3 pax/m^2 we would add 360 standing passengers (for a total of 637 passengers per convoy), and so on.

In this second instance, the RS MIP model sizes of the complete model (RSM) and the reduced model (Reduced RSM) calculated by the Cplex presolver are shown in Table 7. The model is larger than that of line C5.

Table 8
Lines C3–C4 RS model solutions.

$p_{a,l}^{3-4}, p_{a,l}^{4-10}$	#C	TSOC	EMOC	PEC	#CC	OI	ST
0.5,3	62	87170.89	3951.18	2751	28	26.60	171.92
1,5	63	87837.52	4090.83	2177	34	26.36	98.03
3,6	62	88094.56	4090.83	2462	34	26.27	142.60

Table 9
Comparing the non-robust and robust solutions for lines C3–C4.

Case	TSOC	EMOC	#EMRH	PEC	#CC	#CCRH	OI	ST
NoRob	87837.52	4090.83	20	2177	34	18	26.37	98.03
Rob	88085.11	4044.18	14	2177	32	16	26.34	204.48

Table 10
The current RS solution operated by RENFE for lines C3 and C4.

	TSOC	EMOC	PEC	#CC	OI
Current solution	136633.86	6083.88	1550	0	15.20

As for line C5, different solutions can be obtained by varying the penalties for passengers in excess, ($p_{a,l}^{3-4}, p_{a,l}^{4-10}$). These results are summarized in Table 8, which is composed of the same rows and columns that appear in Table 3.

As we increase the excess passenger penalties, we can see that the TSOC increases because more capacity must be offered to meet the same demand with a higher quality of service. However, these costs are slightly lower in every presented solution than in RENFE's current solution (Table 10). PEC are greater than in RENFE's current solution but, as we have shown before, this increment is mainly due to the fact that we are penalizing passengers between 3 and 3.5 pax/m². However, these passengers are more comfortable because they are under the desired limit density of 3.5 pax/m², and thus the quality of service is not degrading. The ST is greater than that used for line C5 because the complexity of the model is greater for this study case.

As we have stated above, robustness may be introduced through different approaches, including avoiding dangerous empty movements and composition changes at rush hours. To illustrate this effect, two different case studies are shown, as for line C5: one in which no robustness is introduced (NoRob) and a second in which robustness is introduced (Rob). The computational results are shown in Table 9.

We can see the differences between the non-robust and robust cases in Table 9. For the robust case, TSOC are greater because more train services are set to double composition; this cost may represent the robustness cost. In the robust case, EMOC are lowered because robustness is achieved by penalizing them. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is strongly reduced. In a similar way, the robustness introduced by composition changes is shown in columns #CC and #CCRH, representing the number of composition changes in the planning period and the composition changes during rush hours, respectively.

Again, PEC always remained similar or equal to the non-robust case. Hence, the price of robustness does not arise from passengers. However, the ST was always greater in the robust case.

In Table 10, we compare the obtained solutions with the actual solution provided by RENFE. As can be seen from the results, the proposed solutions in Tables 8 and 9 improve the current solution, and even after introducing robustness into the model, costs are notably lowered.

6. Conclusions

The rolling stock model presented in this paper is a new approach in the emerging area of urban Rapid Transit network optimization. The results are satisfactory, because in addition to commercial train services, they also account for empty movements, adequate allocation of material in the depots, the optimal convoy mix forming the trains, and rotation times. Another aspect considered in the model is the simultaneous inclusion of all daily rush hours.

Robustness was introduced into the model through a number of different approaches. First, composition changes are penalized depending on the depot stations and time periods. At congested depot stations, it might be very difficult to perform a composition change. Similarly, empty movements are penalized because they use the same infrastructure as commercial train services.

Demand changes due to possible delays were considered. Critical train services are introduced to try to ensure that enough capacity exists under conditions of great demand and short delays. This allows for delayed train service during rush hours to meet passenger demand at congested stations.

The results obtained in the network tests were satisfactory: operating costs were lowered while a high level of service quality for passengers was maintained and robust plans for network operation were provided. Moreover, the time needed to obtain these plans was reduced from the current system of manual planning under great time pressure. In addition, the possibility of analyzing several scenarios rather than just one is considered quite useful. For example, by varying penalties for excess passengers, we can obtain different solutions providing different qualities of passenger service (i.e., the quality of service is parameterized using the mentioned penalties).

Although we have solved case studies in a modular way, as is done in real-life, we have checked that the model sizes obtained are similar to, or even greater than, those in the related literature. In addition, we show how the computational times are sufficiently low to apply this approach in real-life. This solution method may be applied to the integration of two modules of the network. However, it makes no sense to do so because the unique shared variable is the capacity, and this is currently determined during timetable planning by the infrastructure manager.

Future research may integrate the assignment of rolling stock and routing problems, determining train sequences at the same time as the train services' materials and compositions are chosen. Another approach may be to integrate the timetable design and RS assignment problems. This would allow planners to determine potential train commercial movements through the physical network at the same time as their departure times, materials and compositions; for this purpose, we could think of competitive modules competing for capacity in shared depot stations.

Another area of future research is the recoverability problem. When an incident disturbs the normal operation of the network, a new recovery plan must be designed. Canceling and designing new services and changing departure times from existing train services could be possible recovery actions, among others. In this way, recoverable robustness may be introduced (i.e., designing robust rolling stock assignment with recovery plans for some common incidents during network operation).

Finally, an even more ambitious and difficult topic for future research would be the integration of all phases of the planning process (i.e., the planning of timetables, RS, routing and crew scheduling). Such an ideal system, of which we should not lose sight, is the final goal and would take into account every aspect and level in the planning process.

Acknowledgments

This research was supported by project Grant PT2007-003-08CCPP from the "Ministerio de Fomento, Spain" and by the

“Ministerio de Ciencia e Innovación, Spain” for project TRA2008-06782-C02-01. Special thanks to Ángel Luis Matesanz and Isaac Musat from “RENFE Cercanías Madrid” for the data recollection and calibration. The authors also wish to thank to the anonymous reviewers for their helpful comments on this paper.

References

- [1] Schrijver A. Minimum circulation of railway stock. *CWI Quarterly* 1993;6: 205–17. Center for Mathematics and Computer Science, Amsterdam, The Netherlands.
- [2] Alfieri A, Groot R, Kroon LG, Schrijver A. Efficient circulation of railway rolling stock. *Transportation Science* 2006;40:378–91.
- [3] Fioole PJ, Kroon LG, Maróti G, Schrijver A. A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research* 2006;174(2):1281–97.
- [4] Maróti G. Operations research models for railway rolling stock planning. PhD thesis, Eindhoven, Technische Universiteit Eindhoven; 2006.
- [5] Ben-Khedher N, Kintanar J, Queille C, Stripling W. Schedule optimization at SNCF: from conception to day of departure. *Interfaces* 1998;28:6–23.
- [6] Cordeau JF, Soumis F, Desrosiers J. A benders decomposition approach for the locomotive and car assignment problem. *Transportation Science* 2000;34: 133–49.
- [7] Cordeau JF, Desaulniers G, Lingaya N, Soumis F, Desrosiers J. Simultaneous locomotive and car assignment at VIA rail Canada. *Transportation Research B* 2001;35:767–87.
- [8] Mellouli T, Suhl L. Rotation planning of locomotive and carriage groups with shared capacities. In: Geraets F, Kroon LG, Schöbel A, Wagner D, Zaroliagis C, editors. *Algorithmic methods for railway optimization*. Berlin, Germany: Springer; 2007.
- [9] Marín A, Cadarso L. ARRIVAL WP-4 TR-0239. An approach to rapid transit rolling stock assignment, <<http://arrival.cti.gr/uploads/Documents.0239/ARRIVAL-TR-0239.pdf>>; 2009 [Accessed on July 2010].
- [10] Brucker P, Hurink J, Rolfes T. Routing of railway carriages: a case study. *Osnabrücker Schriften zur Mathematik, Reihe P, Heft 205*, 1998.
- [11] Nielsen LK, Kroon L, Maróti G. Short-term rolling stock rescheduling. Large-scale optimization: robustness, online and offline issues 3rd (Final) ARRIVAL Review Meeting, May 2009, Patras, Grecia. <<http://arrival.cti.gr/uploads/presentations/Nielsen-Short-Term-Rolling-Stock-Rescheduling.pdf>> [Accessed on July 2010].
- [12] Marín A, Cadarso L. ARRIVAL WP-4 TR-0240. The rapid transit routing, 2009. <<http://arrival.cti.gr/uploads/Documents.0240/ARRIVAL-TR-0240.pdf>> [Accessed on July 2010].
- [13] Cadarso L, Marín A. Robust routing of rapid transit rolling stock. *Public Transport* 2010;2:51–68.
- [14] Freling R, Lentink RM, Kroon LG, Huisman D. Shunting of passenger train units in a railway station. *Transportation Science* 2005;39(2):261–72.
- [15] Lentink RM. Algorithmic decision support for shunt planning. PhD thesis, Erasmus University Rotterdam, The Netherlands; 2006.
- [16] Van den Broek JJJ, Kroon LG. A capacity test for shunting movements. In: Geraets F, Kroon LG, Schöbel A, Wagner D, Zaroliagis C, editors. *Algorithmic methods for railway optimization*. Berlin, Germany: Springer; 2007.
- [17] de Almeida D, Trouve F, Hebben B. A decision aid module for resources rescheduling when managing disturbances. Technical Report, Presented at the World Congress on Railway Research; 2003.
- [18] Liebchen C, Möhring R, Lübbecke M, Stiller S. ARRIVAL Deliverable D1.2. new theoretical notion of the prices of robustness and recoverability, <<http://arrival.cti.gr/uploads/deliverables1/ARRIVAL-Del-D1.2.pdf>>; 2007 [Accessed on July 2010].
- [19] Cacchiani V, Caprara A, Galli L, Kroon L, Maróti G, Toth P. ARRIVAL WP-4 TR-121. Price of recoverability for railway rolling stock planning, <<http://arrival.cti.gr/uploads/Documents.0121/ARRIVAL-TR-0121.pdf>>; 2008 [Accessed on July 2010].
- [20] Cacchiani V, Caprara A, Galli L, Kroon L, Maróti G, Toth P. ARRIVAL WP-4 TR-186. Recoverable robustness for railway rolling stock planning, <<http://arrival.cti.gr/uploads/Documents.0186/ARRIVAL-TR-0186.pdf>>; 2009 [Accessed on July 2010].
- [21] Daskin M, Panayotopoulos N. A Lagrangian relaxation approach to assigning aircraft to routes in hub and spoke networks. *Transportation Science* 1989;23: 91–9.
- [22] Abara J. Applying integer linear programming to the fleet assignment problem. *Interfaces* 1989;19(4):20–8.
- [23] Hane C, Barnhart C, Johnson E, Martsen R, Nemhauser G, Sigismondi G. The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming* 1995;70:211–32.
- [24] Subramanian R, Scheff RP, Quillian JD, Wiper DS, Marsten RE. Coldstart: fleet assignments at delta air lines. *Interfaces* 1994;24(1):104–20.